Management of Pornography-seeking in an Online Dermatology Atlas: Adventures in the Skin Trade[1]

Christoph U. Lehmann MD[a], Bernard A. Cohen MD[b], George R. Kim, MD

[a]Division of Health Sciences Informatics and [b]Department of Dermatology, Johns Hopkins School of Medicine, Baltimore, Maryland USA

Abstract

The escalating competition between online pornography - seeking and disseminating behaviors and technologies that attempt to reduce them creates technical, semantic and legal barriers to the legitimate discussion of and education about sensitive health issues involving sexuality, anatomy and pathology, especially when image-based knowledge is used. The effects of this competition on the use and management of an online dermatology atlas are described with a discussion on the importance of anticipating, addressing and controlling this problem while developing and maintaining image-based digital libraries and other e-Health applications.

Introduction

The prevalence of online pornography-seeking and disseminating behaviors and the development of technical methods to reduce them present ethical, legal and technical barriers to legitimate discussion and education about sensitive health issues such as sexuality, breastfeeding, child maltreatment and the anatomic manifestations of disease. While approximately 4.5% of search engine queries relate to bona fide health topics[2], it has been is estimated that 20-25% of all Internet searches relate to pornography[3]. Technical approaches to block pornographic material have been primarily (English) text-based, and have sometimes resulted in inappropriate and inconvenient censorship on specific health topics[4], and while algorithmic improvements[5] have increased filter precision, they are not perfect[6]. Filtering of visual media (image, video) using approaches based on image content and color[7] have been in development but are not yet widely used.

The increasing availability of collaborative online image libraries of high quality, integrity and educational value in pediatrics[8], surgery[9], ophthalmology[10], anatomy[11], histology[12, 13], pathology[14], radiology[15, 16] and dermatology[17] has unfortunately also increased the potential for misuse of health information. Negative effects of the inappropriate use of such archived information and digital image libraries are three-fold: 1) pornography seeking (and dissemination) behaviors reduce the impact and integrity of legitimate health information through its misuse and misrepresentation, 2) the resultant "noise" from such behaviors decreases the prevalence of legitimate information, making it difficult to find, and 3) filtering efforts block access to legitimate health information[18]. Developers and curators of online health libraries must be aware of and sensitive to these issues to protect the integrity of their collections and the services they provide.

This study describes our experience with the public use of a curated, online dermatologic image library, focusing on the management of its misuse, and the perils associated with the publication of this collection over a period of four years.

Methods

The health image repository studied is DermAtlas[16]. Released in December 2000 by the Johns Hopkins University School of Medicine, it is the largest online dermatology image library, providing open access to high quality clinical and histological images with associated text about physical findings and related information. DermAtlas currently contains 7,300+ images (of which 3,900 relate to pediatrics) contributed by health professionals from around the world. Images may be located via a search engine using entered query terms for diagnosis, disease category, body site, patient age and gender. The database and image library are managed closely by its curators (CUL, BAC) to assure the availability and integrity of its content and sources. DermAtlas has more than 28,000 unique visitors daily, with an average of 5.3 pages viewed per visit.

As part of ongoing assessment of user information needs, all queries are recorded in a database (with their source (client Internet Protocol (IP) address) and time of entry) and periodically analyzed to guide content development and expansion. Other data reviewed in this study included server logs of client IP addresses, referrer information and time stamps.

One year post-deployment, a marked increase was noted in both the frequency of searches for images containing genital sites and of retrieval of such images. Review and analysis of searches performed

in DermAtlas were conducted in conjunction with an analysis of DermAtlas content. Statistical methods used include Chi-Square test.

Results

Of 3,664,191 queries entered by users during the time period 10/19/2004 to 3/5/2005, 10.9% contained one of 108 defined anatomic sites. Each anatomic site had been searched at least once (Range: 6 ("conjunctiva, palpebral") - 26969 ("penis"), median = 2045.5, mean = 3694). Of these, 37% were for one of 14 genital sites. Of the ten most frequently entered anatomic sites, seven were of genital sites (Figure 1).

Of all queries, 62% were for one of 1139 diagnoses. The number of searches per diagnosis ranged from 4 (eosinophilic folliculitis) to 117,828 (Herpes simplex virus infection) (Median 644, Mean 1,992,). Of those queries for diagnoses, 12.5% were linked with images of a genital location (such as anogenital warts or diaper dermatitis).

Of all queries, 10,307 (2.8%) specified both an anatomic site AND an age range (such as 2-6 months). Of these queries, 3,442 (33.4%) specified a genital site, 5,840 (56.7%) specified children and 4,467 (43.3%) specified adults. Of those specifying genital sites, 2,499 (72.6%) also specified children AND a genital site, while 954 (27.7%) also specified adults ($\chi^2 = 513.7$, N=10,307, df=1, $p < 0.001$) (Table 1).

	Child	Adult
Genital	2488	954
Non-Genital	3352	3513

Table 1: Queries for Age and Anatomic Sites

In February 2005, 10,000 consecutive free-text queries were analyzed and classified according to their retrieval of genital images (yes/no). A cross-tabulation generated 3,919 unique query entries (including misspellings of the same concept). Many queries were classified as clearly inappropriate (Examples: "large penis", "sex", "girl vagina", "penile affections", "erection", "pussy"). A total of 1,208 (12.1%) of the queries retrieved images of a genital region. Of all DermAtlas images, 5.5% contain genital anatomic sites.

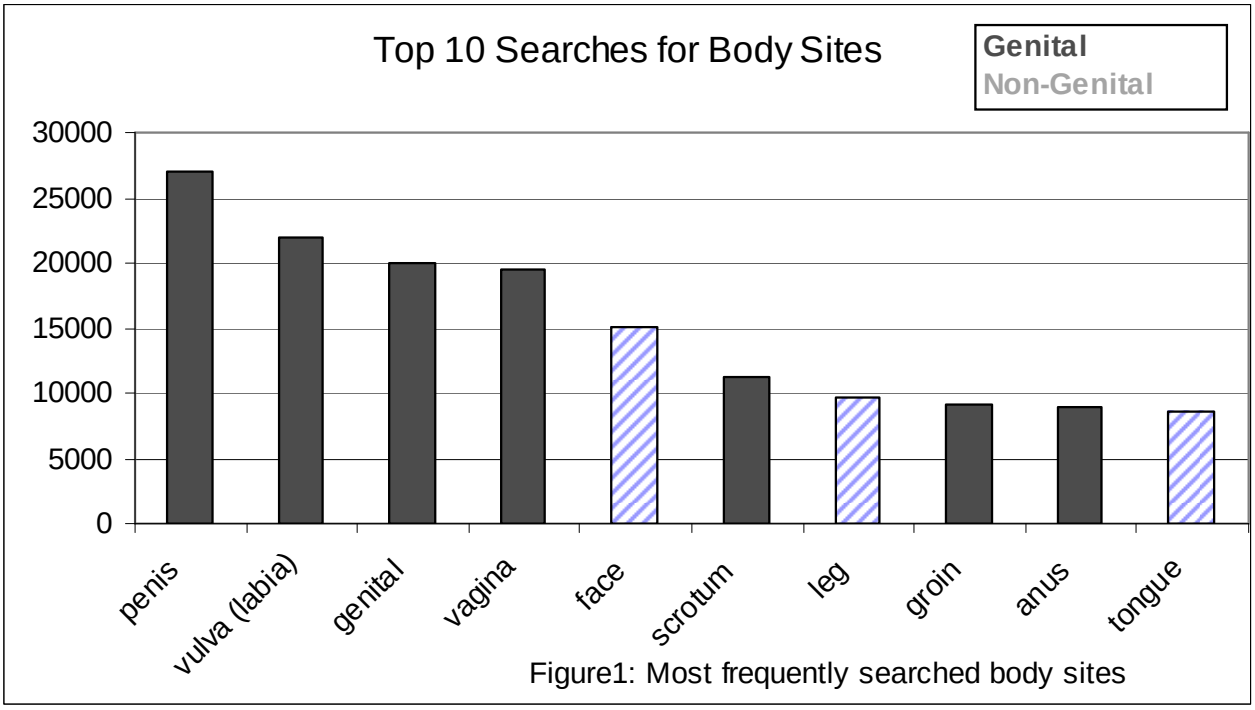

Figure1: Most frequently searched body sites

Discussion

Developers of image-based digital libraries (DLs) create them for education and reference. While DLs are intended for multiple uses, their content may be used in unexpected, undesired and sometimes unforeseen ways that compromise their integrity. In the case of DermAtlas, evidence from user queries and log files indicated that this image library was (and is) being used inappropriately.

Such misuse undermines not only the intent of the curators, but also the goodwill of contributors to the library and the patients who have consented to have their photographs archived. For continued survival of

DLs, the maintenance of their integrity and of the collection process used must be preserved.

Detecting and curbing inappropriate access

What was initially a tool to guide DermAtlas content development has became an instrument to curb its inappropriate use. Analysis of archived user queries and log files help to determine characteristics of misuse, which in turn, helps to develop rules to filter inappropriate queries. Due to changing usage patterns, this is a continuously evolving process. Initially, filters based on Internet Protocol (IP) addresses or on Websites that are sources of frequent queries for genital images blocked legitimate requests from practitioners who frequently use them to educate and reassure patients, from users at shared workstations in libraries and from users whose requests are routed through proxy servers. Three major entities that channel requests through proxy servers include the National Health Service (NHS) in Great Britain, military services in the US and one Internet Service Provider (ISP) in Canada, which have resulted in a large number of access denials to their subscribers. In case of the NHS, the filter was suspended, but in light of increasing reports of child pornography possession by dentists and physicians in Great Britain[19], this decision may require reconsideration.

The current procedure to curb inappropriate use of DermAtlas includes monitoring and logging of IP addresses associated with more than a specified number of queries for genital sites or images. Users from these logged IP addresses may see thumbnails (small images) but are restricted from retrieving full-size images of genital sites. These users must consult DermAtlas curators for passwords to regain access. To date, there are 9,338 logged IP addresses used in this filter. To allow Internet robots ("spiders") to follow predefined body site queries and index them, known IP addresses of spiders are exempt from this restriction.

Monitoring also includes investigation of surges in DermAtlas use. In the summer of 2003, analysis of a significant increase in uses for a single day revealed direct linkage to genital images from an identified pornography/fetish site. Analysis for February 2005 revealed that of the top 43 referring sites, nine (21%) were from pornographic/fetish sites. Of actual requests, these sites were the source of only 14.3% of the 141,285 referrals (Figure 2). To reduce referrals from such inappropriate sites, referring pages are analyzed. If a referring page is on a list (which is updated bimonthly), referrals from that source are directed to a denial page informing the user that DermAtlas does not desire traffic from the referring site, and the user's IP address is recorded.

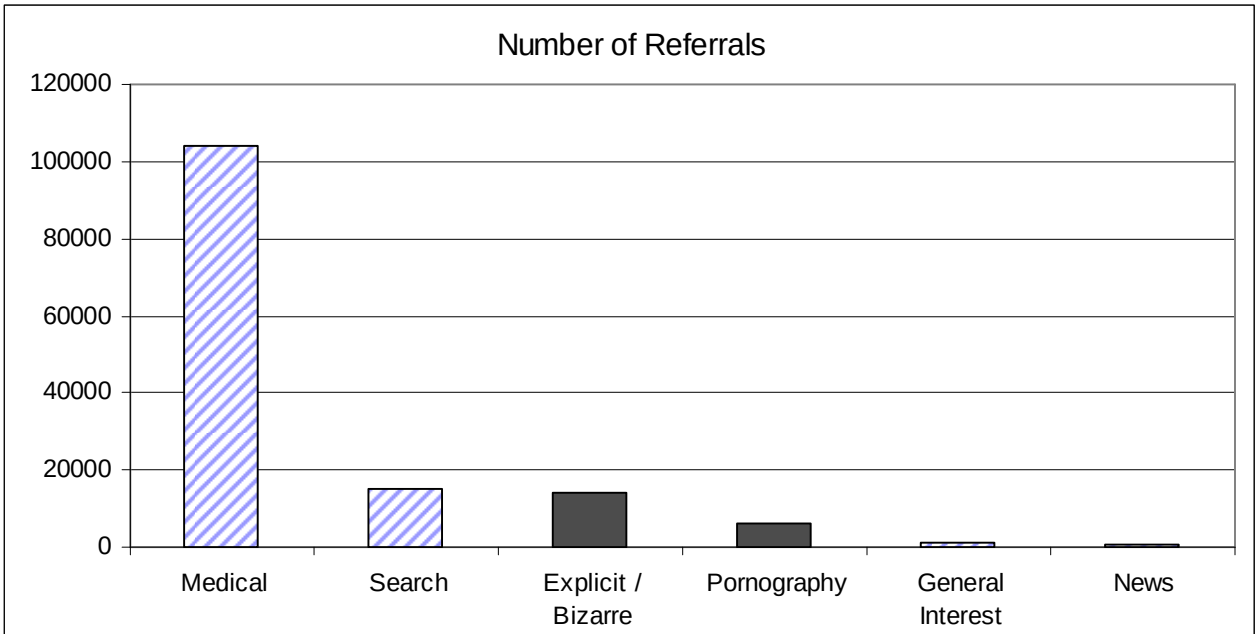

Figure 2: Number of referrals from the 43 most frequent referrers in February 2005 by referral site content

Beyond the Digital Library

We are not aware of any other studies describing use of online clinical digital image libraries or if use

reflects the intent of developers. The potentials of using the information and knowledge contained in any library include the possibility of misuse. While technologies such as the World Wide Web extend the power and distribution of multimedia libraries, they also enable misuse of technology and information. Such misuses of health information carry their own risks and burdens for DL curators including legal challenges, ethical concerns and drains on resources (processing power, bandwidth) of which developers and curators of resources must be aware.

Beyond the technical points discussed in this paper are legal issues related to maintaining the integrity of DLs. Curators of DLs that use case materials from patients must be aware of personal health identifiers (PHI) that may require consideration of adherence to HIPAA privacy rules. In many cases, PHI may have been already removed from stored information, but the needs and expectations of patients who give consent to have their images published in this highly accessible medium must also be honored. Formalized procedures for DLs to assure appropriate use of information, its tracking and protection against its misuse are basic requirements for developing and maintaining the integrity of DLs. However, no universal standards for such procedures exist.

The legal implications of pornography, in particular child pornography[20], for which possession in any format carries significant consequences, must be kept in mind. It is essential that users and contributors (providers and patients) understand the need to protect the integrity of DermAtlas as a source of health information for educational use by health care providers and patients. Even with best intentions, attempts to curb pornography-seeking behaviors may be misinterpreted as accusations of inappropriate access. Filters and rules for blocking are not perfect and there will be false positive blocks, which may offend legitimate users who are denied access. Conversely, false negative (failed) blocks allow continued misuse. Prevention of false negative blocks requires continuous vigilance and adjustment of filters and strategies based on changing usage patterns.

In 2004, a report by two individuals to the National Center for Missing and Exploited Children led to a formal complaint and investigation by the local police of DermAtlas as child pornography. Consultation with institutional lawyers reinforced a prior establishment of DermAtlas' role as an educational tool (similar to a medical text book) for which neither state nor federal child pornography laws apply. Shortly after resolution of the investigation, DermAtlas developers received an email message trying to extort money in return for 'not reporting your posting of child pornography on the Web'. The institutional legal team quickly resolved the issue, and no further attempts at extortion have been reported.

Another issue that DL curators must face in the effective provision of information for education and eHealth services on sensitive topics is the barrier formed by language and literacy problems. Beyond language barriers (DermAtlas is written in Englsh) are cultural and health literacy problems. One method used to curb inappropriate use is to disallow colloquial or slang terms in digital library entries or search engines. However, for some patients, these terms may be the only means by which communication on a topic is initiated. The role of DLs in spectrum of patient care and communication and the relationships among patients, providers and the health system has yet to be explored.

Librarians have long struggled with the challenges of misuse of library computers by patrons to view pornography and the exposure of such materials to other patrons. Congress had enacted the Children's Internet Protection Act (CIPA), which forced libraries to install pornography filters in order to receive federal funds, but the Supreme Court, because of violations of the first amendment, struck down CIPA.[21] Server-side restrictions similar to the described user-filters may also pose constitutional questions for publicly-funded DLs.

Conclusions

Developers of online clinical image libraries must be aware of issues beyond the development and curation of content. The problems and consequences of misuse of the information resources they create and maintain must be anticipated, addressed and prevented if their integrity is to be maintained. Pornography-seeking behaviors and technologies that attempt to reduce them may have negative effects on the accessibility and use of health information, the legitimate discussion of and education about sensitive health issues and on e-Health in general.

References:

[1] With apologies to Dylan Thomas.

[2] Eysenbach G, Kohler Ch. What is the prevalence of health-related searches on the World Wide Web? Qualitative and quantitative analysis of search engine queries on the internet. AMIA Annu Symp Proc. 2003;:225-9.

[3] Phillipov G, Phillips PJ. Frequency of health-related search terms on the Internet. JAMA. 2003 Nov 5;290(17):2258-9.

[4] Beeson A. Top Ten Threats to Civil Liberties in Cyberspace. Human Rights Vol. 23, No. 2 Spring 1996. URL: http://www.abanet.org/irr/hr/hrtopten.html [Last accessed 6 March 2005].

[5] Su GY, Li JH, Ma YH, Li SH. Improving the precision of the keyword-matching pornographic text filtering method using a hybrid model. J Zhejiang Univ Sci. 2004 Sep;5(9):1106-13.

[6] Fallows D. Spam: How It Is Hurting Email and Degrading Life on the Internet. Pew Internet and American Life Project. October 22, 2003. URL: http://www.pewinternet.org/pdfs/PIP_Spam_Report.pdf [Last accessed 6 March 2005].

[7] Forsyth D, Categorization of Images, in Technical, Business, and Legal Dimensions of Protecting Children from Pornography on the Internet: Proceedings of a Workshop. The National Academies Press, 2002.

[8] Pediatric Picture Atlas. http://www.rainbowpediatrics.net/cgi-bin/album.pl?album=PPA [last accessed 3/6/2005]

[9] Online Atlas of Surgery. http://www.wfubmc.edu/surg-sci/atlas/atlas.html [last accessed 3/6/2005]

[10] Eyeatlas http://www.eyeatlas.com/ [last accessed 3/6/2005]

[11] Atlas of the body. http://www.ama-assn.org/ama/pub/category/7140.html [last accessed 3/6/2005]

[12] Atlas of Microscopic Anatomy: A Functional Approach: Companion to Histology and Neuroanatomy: Second Edition. http://www.vh.org/adult/provider/anatomy/MicroscopicAnatomy/MicroscopicAnatomy.html#TOC [last accessed 3/6/2005]

[13] Atlas of Histology. http://www.med.uiuc.edu/histo/large/atlas/ [last accessed 3/6/2005]

[14] The Urbana Atlas of Pathology. http://www.med.uiuc.edu/pathatlasf/titlepage.html [last accessed 3/6/2005]

[15] Radiology Atlas. http://ect.downstate.edu/courseware/rad_atlas/ [last accessed 3/6/2005]

[16] Radiology Atlas Images. http://anatomy.uams.edu/anatomyhtml/xrays/rad_atlas.html [last accessed 3/6/2005]

[17] DermAtlas. http://www.dermatlas.org [last accessed 3/6/2005]

[18] Richardson CR Resnick PJ, Hansen DL, Derry HA, Rideout VJ. Does Pornography-Blocking Software Block Access to Health Information on the Internet? JAMA. 2002;288:2887-2894.

[19] White C. GMC sees rise in doctors charged with accessing child pornography websites. BMJ. 2004 Apr 24; 328(7446):973.

[20] "Sexual Exploitation and Other Abuse of Children," Title 18 U.S. Code, Part I, Chapter 110, 2003 ed.

[21] United States vs. American Library Association. United States Supreme Court Decision. Decided June 23, 2003 http://www.supremecourtus.gov/opinions/02pdf/02-361.pdf [last accessed 7/10/2005]

Corresponding author:
Christoph U. Lehmann, MD
Division of Health Sciences Informatics
Johns Hopkins University SOM
Nelson 2-133
600 North Wolfe Street
Baltimore, Maryland 21287-3200